PCT

W[…]D INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 38/00, 39/395, 48/00, C07H 21/04, C07K 14/435, 14/705, 16/00, 16/18, 16/28, C12N 15/01, 15/11, 15/12, 15/63

A1

(11) International Publication Number: **WO 98/53840**

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/10791

(22) International Filing Date: 27 May 1998 (27.05.98)

(30) Priority Data:
60/047,572 27 May 1997 (27.05.97) US

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KIKLY, Kristine, Kay [US/US]; 499 Limerick Center Road, Linfield, PA 19468 (US). ERICKSON-MILLER, Connie, Lynn [US/US]; 608 Brainerd Place, Exton, PA 19341 (US).

(74) Agents: ANDERSEN, Robert, L. et al.; Ratner & Prestia, Suite 301, One Westlakes, Berwyn, P.O. Box 980, Valley Forge, PA 19482–0980 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SIALOADHESIN FAMILY 4 cDNA

(57) Abstract

The SAF–4 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing SAF–4 polypeptides and polynucleotides in therapy, and diagnostic assays for such.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Sialoadhesin Family 4 cDNA

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterize further genes and their related polypeptides/proteins, as targets for drug discovery.

Summary of the Invention

The present invention relates to SAF-4, in particular SAF-4 polypeptides and SAF-4 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amylotrophic lateral sclerosis, head injury damage, and other neurological abnormalities, septic shock, sepsis, stroke, osteoporosis, osteoarthritis, ischemia reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischemic injury, myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes and other hematologic abnormalities, aplastic anemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and

antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with SAF-4 imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate SAF-4 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to SAF-4 polypeptides. Such peptides include isolated polypetides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include the polypeptide of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides of the present invention are believed to be members of the Sialoadhesin family of polypeptides. They are therefore of interest because the sialoadhesin family of proteins, sialoadhesin, CD33, CD22 and myelin-associated glycoprotein (MAG), are utilized as cellular interaction molecules. They bind specific carbohydrates in a sialic acid-dependent manner on target cells. The extracellular domain is made up of various numbers of immunoglobulin-like domains of the V-like and C2-like subtypes and the intracellular portion has no known homology to any signalling motifs. Sialoadhesin expression is restricted to macrophages, it has 17 Ig-like domains and the specific recognition sequence on target cells is Neu5Aca2,3Galb13GalNAc. Known target cells are developing myeloid cells in the bone marrow and lymphocytes in the spleen and lymph node (Crocker, P. R., et al. EMBO J,

1994, 13:4490-4503). CD22 is expressed only on B cells and has a and b isoforms with 5 and 7 Ig-like domains, respectively. CD22 is known to bind T cells, B cells, monocytes, granulocytes and erythrocytes by recognizing Neu5Aca2,6Galb1,4Glc(NAc) in N-linked glycans (Crocker, P. R., et al. EMBO J, 1994, 13:4490-4503; Stamenkovic, I. and Seed, B. Nature, 1990, 345:74-77; Wilson, G. L., et al. J Exp Med, 1991, 173:137-146). Myelin-associated glycoprotein (MAG) is expressed by Schwann cells in the peripheral nervous system and oligodendrocytes in the central nervous system and is thought to participate in the cell adhesion to axons. MAG has two alternatively spliced variants, large MAG (L-MAG) and small MAG (S-MAG) which are expressed either during embryonic development or in the adult, respectively. The alternative splicing results in the expression of the same extracellular domains but distinct intracellular domains (Pedraza, L. et al., JCB, 1990, 111:2651-2661).

CD33 is most relevant to SAF-4 because they are the most closely related of all the family members. CD33 is normally expressed on the developing myelomonocytic lineage. It is absent on early stem cells but is present on colony-forming units for granulocytes, erythrocytes, monocytes, and megakaryocytes (CFU-GEMM) and progenitors of granulocytes and mononuclear phagocytes (CFU-GM). It is downregulated by mature granulocytes but retained by mature monocytes and macrophages (Andrews, R. G., et al., Blood, 1983, 62:124; Griffin, J. D., et al., Leuk Res 1984, 8:521). CD33 has two Ig-like domains and prefers to bind targets expressing NeuAca2,3Gal in N- and O-linked glycans. It maps to chromosome 19q13.1-13.3, closely linking it in the genome with MAG and CD22 (Freeman, S. D., et al., Blood, 1995, 85:2005-2012).

CD33 has also been found to be expressed on about 85% of leukemic myeloblasts in patients with acute myelogenous leukemia (AML) and is frequently used to differentiate AML from acute lymphoblastic leukemia (ALL). Monoclonal antibodies to CD33 have been used therapeutically to purge residual myeloblasts from autologous bone marrow grafts ex vivo for the treatment of AML (Robertson, M. J., et al., Blood, 1992 79:2229-2236). More recently, humanized monoclonal antibodies to CD33 have undergone evaluation in vivo for the treatment of AML (Caron, P. C., et al., Blood, 1994, 83:1760-1768). These properties are hereinafter referred to as "SAF-4 activity" or "SAF-4 polypeptide activity" or "biological activity of SAF-4". Also included amongst these activities are antigenic and

immunogenic activities of said SAF-4 polypeptides, in particular the antigenic and immunogenic activities of the polypeptide of SEQ ID NO:2. Preferably, a polypeptide of the present invention exhibits at least one biological activity of SAF-4.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The present invention also includes include variants of the aforementioned polypetides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to SAF-4 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1.

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequence of SEQ ID NO:1 shows homology with CD33 (Simmons, D., and Seed, B., JI 141:2797-2800, 1988). The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 51 to 1970) encoding a polypeptide of 639 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is structurally related to other proteins of the Sialoadhesin family, having homology and/or structural similarity withCD33 (Simmons, D., and Seed, B., JI 141:2797-2800, 1988).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and

polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one SAF-4 activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1 and SEQ ID NO:2.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

(a) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(b) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(c) the polynucleotide of SEQ ID NO:3; or

(d) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;

as well as the polynucleotide of SEQ ID NO:3.

The present invention further provides for a polypeptide which:

(a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity,

most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(c) comprises the amino acid of SEQ ID NO:4; and

(d) is the polypeptide of SEQ ID NO:4;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

The nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognized by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded therefrom are therefore subject to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:3 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human primary dendritic cells, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al. Science* (1991) 252:1651-1656; Adams, M.D. *et al.*, *Nature*, (1992) *355*:632-634; Adams, M.D., *et al.*, *Nature* (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention,

the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42^{o}C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65^{o}C.

Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerization reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™' technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are

genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY*

MANUAL (*supra*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterized by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to

the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled SAF-4 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc Natl Acad Sci USA* (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising SAF-4 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the Diseases through detection of mutation in the SAF-4 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays and flow cytometric analysis.

Thus in another aspect, the present invention relates to a diagonostic kit which comprises:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof ;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amylotrophic lateral sclerosis, head injury damage, and other neurological abnormalities, septic shock, sepsis, stroke, osteoporosis, osteoarthritis, ischemia reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischemic injury, myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes and other hematologic abnormalities, aplastic anemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the

affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others. Antibodies against SAF-4 polypeptides may also be employed to subcharacterize cell populations during hematopoietic development, as a diagnostic marker to distinguish between different forms of cancer, to purge bone marrow ex vivo of cancer cells expressing SAF-4, as a tool to aid in the ex vivo expansion (proliferation and/or differentiation) of hematopoietic progenitor cells expressing SAF-4, as a stimulus in vivo for stem cell mobilization into the periphery, and as an in vivo chemoprotective agent.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. In another approach, soluble forms of SAF-4 polypeptides still capable of binding the ligand in competition with endogenous SAF-4 may be administered. Typical embodiments of such competitors comprise fragments of the SAF-4 polypeptide. One example is using the extracellular domain of SAF-4 fused to a human immunoglobulin Fc region which can then be employed to treat cancer, inflammation, autoimmunity and allergy, among others. SAF-4/Fc polypeptides may also be employed to purge bone marrow ex vivo of cancer cells expressing SAF-4 ligands, as a tool to aid in the ex vivo expansion (proliferation and/or differentiation) of hematopoietic progenitor cells expressing SAF-4 ligands, as a stimulus in vivo for stem cell mobilization into the periphery, and as an in vivo chemoprotective agent. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, *Current Protocols in Immunology* 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polpypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring SAF-4 activity in the mixture, and comparing the SAF-4 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and SAF-4 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is

labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesizing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an interactive process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amylotrophic lateral sclerosis, head injury damage, and other neurological abnormalities, septic shock, sepsis, stroke, osteoporosis, osteoarthritis, ischemia reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischemic injury, myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes and other hematologic abnormalities, aplastic anemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections, related to either an excess of, or an under-expression of, SAF-4 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the SAF-4 polypeptide.

In still another approach, expression of the gene encoding endogenous SAF-4 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, *J Neurochem* (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al., Nucleic Acids Res* (1979) 6:3073; Cooney *et al., Science* (1988) 241:456; Dervan *et al., Science* (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of SAF-4 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of SAF-4 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include,

but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCC. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus,

"polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York,

1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer*

Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol. 215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol. 215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO: 1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$\mathbf{n_n \leq x_n - (x_n \bullet y)},$$

wherein $\mathbf{n_n}$ is the number of nucleotide alterations, $\mathbf{x_n}$ is the total number of nucleotides in SEQ ID NO:1, **y** is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and $\bullet$ is the symbol for the multiplication operator, and wherein any non-integer product of $\mathbf{x_n}$ and **y** is rounded down to the nearest integer prior to subtracting it from $\mathbf{x_n}$. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of amino acid alterations, x_n is the total number of amino acids in SEQ ID NO:2, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., • is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n.

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO: 2 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid

alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and • is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

SEQUENCE INFORMATION

SEQ ID NO:1

TCCGAGAGAACAGGCCTGTCTCAGGCAGGCCCTGCGCCTCCTATGCGGAGATGCTACTGCCACTGCTGCTGTCC
TCGCTGCTGGGCGGGTCCCAGGCTATGGATGGGAGATTCTGGATACGAGTGCAGGAGTCAGTGATGGTGCCGGA
GGGCCTGTGCATCTCTGTGCCCTGCTCTTTCTCCTACCCCCGACAAGACTGGACAGGGTCTACCCCAGCTTATG
GCTACTGGTTCAAAGCAGTGACTGAGACAACCAAGGGTGCTCCTGTGGCCACAAACCACCAGAGTCGAGAGGTG
GAAATGAGCACCCGGGGCCGATTCCAGCTCACTGGGGATCCCGCCAAGGGGAACTGCTCCTTGGTGATCAGAGA
CGCGCAGATGCAGGATGAGTCACAGTACTTCTTTCGGGTGGAGAGAGGAAGCTATGTGAGATATAATTTCATGA
ACGATGGGTTCTTTCTAAAAGTAACAGTGCTCAGCTTCACGCCCAGACCCCAGGACCACAACACCGACCTCACC
TGCCATGTGGACTTCTCCAGAAAGGGTGTGAGCGCACAGAGGACCGTCCGACTCCGTGTGGCCTATGCCCCCAG
AGACCTTGTTATCAGCATTTCACGTGACAACACGCCAGCCCTGGAGCCCCAGCCCCAGGGAAATGTCCCATACC
TGGAAGCCCAAAAAGGCCAGTTCCTGCGGCTCCTCTGTGCTGCTGACAGCCAGCCCCCTGCCACACTGAGCTGG
GTCCTGCAGAACAGAGTCCTCTCCTCGTCCCATCCCTGGGGCCCTAGACCCCTGGGGCTGGAGCTGCCCGGGGT
GAAGGCTGGGGATTCAGGGCGCTACACCTGCCGAGCGGAGAACAGGCTTGGCTCCCAGCAGCGAGCCCTGGACC
TCTCTGTGCAGTATCCTCCAGAGAACCTGAGAGTGATGGTTTCCCAAGCAAACAGGACAGTCCTGGAAAACCTT
GGGAACGGCACGTCTCTCCCAGTACTGGAGGGCCAAAGCCTGTGCCTGGTCTGTGTCACACACAGCAGCCCCCC
AGCCAGGCTGAGCTGGACCCAGAGGGGACAGGTTCTGAGCCCCTCCCAGCCCTCAGACCCCGGGGTCCTGGAGC
TGCCTCGGGTTCAAGTGGAGCACGAAGGAGAGTTCACCTGCCACGCTCGGCACCCACTGGGCTCCCAGCACGTC
TCTCTCAGCCTCTCCGTGCACTACTCCCCGAAGCTGCTGGGCCCCTCCTGCTCCTGGGAGGCTGAGGGTCTGCA
CTGCAGCTGCTCCTCCCAGGCCAGCCCGGCCCCCTCTCTGCGCTGGTGGCTTGGGGAGGAGCTGCTGGAGGGGA
ACAGCAGCCAGGACTCCTTCGAGGTCACCCCCAGCTCAGCCGGGCCCTGGGCCAACAGCTCCCTGAGCCTCCAT
GGAGGGCTCAGCTCCGGCCTCAGGCTCCGCTGTGAGGCCTGGAACGTCCATGGGGCCCAGAGTGGATCCATCCT
GCAGCTGCCAGATAAGAAGGGACTCATCTCAACGGCATTCTCCAACGGAGCGTTTCTGGGAATCGGCATCACGG
CTCTTCTTTTCCTCTGCCTGGCCCTGATCATCATGAAGATTCTACCGAAGAGACGGACTCAGACAGAAACCCCG
AGGCCCAGGTTCTCCCGGCACAGCACGATCCTGGATTACATCAATGTGGTCCCGACGGCTGGCCCCCTGGCTCA
GAAGCGGAATCAGAAAGCCACACCAAACAGTCCTCGGACCCCTCTTCCACCAGGTGCTCCCTCCCCAGAATCAA
AGAAGAACCAGAAAAAGCAGTATCAGTTGCCCAGTTTCCCAGAACCCAAATCATCCACTCAAGCCCCAGAATCC
CAGGAGAGCCAAGAGGAGCTCCATTATGCCACGCTCAACTTCCCAGGCGTCAGACCCAGGCCTGAGGCCCGGAT
GCCCAAGGGCACCCAGGCGGATTATGCAGAAGTCAAGTTCCAATGAGGGTCTCTTAGGCTTTAGGACTGGGACT
TCGGCTAGGGAGGAAGGTAGAGTAAGAGGTTGAAGATAACAGAGTGCAAAGTTTCCTTCTCTCCCTCTCTCTCT
CTCTTTCTCTCTCTCTCTCTTTCTCTCTCTTTAAAAAACATCTGGCCAGGCAGCACAGTGGCTCACGCCTGT
AATCCCAGCACTTTGGGAGGTTGAGGTGGGCAGATCGCCTGAGGTCGGGAGTTCGAGACCAGCCTGGCCAACTT
GGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGGCGCCTGTAATCCTACCTACT
TGGGAAGCTGAGGCAGGAGAATCACTTGAACCTGGGAGACGGAGGTTGCAGTGAGCCAAGATCACACCATTGCA
CGCCAGCCTGGGCAACAAAGCGAGACTCCATCTCAAAAAAAAAATCCTCCAAATGGGTTGGGTGTCTGTAATCC
CAGCACTTTGGGAGGCTAAGGTGGGTGGATTGCTTGAGCCCAGGAGTTCGAGACCAGCCTGGGCAACATGGTGA
AACCCCATCTCTACAAAAAATACAAAACATAGCTGGGCTTGGTGGTGTGTGCCTGTAAGTCCCCAGCTGTCAGA

CATTTAAACCAGAGCAACTCCATCXTGGAATAGGAGGCTGAATAAAATGAGGCTGAGACCTXACTGGGCTGCAT
TCTCCAGACAGTGGAGGCATTCTAAGTCACAGGATGAGACAGGAGGTCCCGTACAAGATACAGGTCATAAAGAC
TTTGCTGATAAAACAGATTGCAGTAAAGAAGCCAACCAAATCCCACCAAAACCAAGTTGGCCACGAGAGTGACC
TCTGGTCGTCCTCACTGCTACACTCCTGACAGCACCATGACAGTTTACAAATGCCATGGCAACATCAGGAAGTT
ACCCGATATGTCCCAAAAGGGGGAGGAATGAATAATCCACCCCTTGTTTAGCAAATAAGCAAGAAATAACCATA
AAAGTGGGCAACCAGCAGCTCTAGGCGCTGCTCTTGTCCTATGGAGTAGCCTTCTTTTTTCCTTTAACTTTCTT
GATAAACTTGCTTTCACCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID NO:2

MLLPLLLSSLLGGSQAMDGRFWIRVQESVMVPEGLCISVPCSFSYPRQDWTGSTPAYGYWFKAVTETTKGAPVA
TNHQSREVEMSTRGRFQLTGDPAKGNCSLVIRDAQMQDESQYFFRVERGSYVRYNFMNDGFFLKVTVLSFTPRP
QDHNTDLTCHVDFSRKGVSAQRTVRLRVAYAPRDLVISISRDNTPALEPQPQGNVPYLEAQKGQFLRLLCAADS
QPPATLSWVLQNRVLSSSHPWGPRPLGLELPGVKAGDSGRYTCRAENRLGSQQRALDLSVQYPPENLRVMVSQA
NRTVLENLGNGTSLPVLEGQSLCLVCVTHSSPPARLSWTQRGQVLSPSQPSDPGVLELPRVQVEHEGEFTCHAR
HPLGSQHVSLSLSVHYSPKLLGPSCSWEAEGLHCSCSSQASPAPSLRWWLGEELLEGNSSQDSFEVTPSSAGPW
ANSSLSLHGGLSSGLRLRCEAWNVHGAQSGSILQLPDKKGLISTAFSNGAFLGIGITALLFLCLALIIMKILPK
RRTQTETPRPRFSRHSTILDYINVVPTAGPLAQKRNQKATPNSPRTPLPPGAPSPESKKNQKKQYQLPSFPEPK
SSTQAPESQESQEELHYATLNFPGVRPRPEARMPKGTQADYAEVKFQ.

SEQ ID NO:3

TCCGAGAGAACAGGCCTGTCTCAGGCAGGCCCTGCGCCTCCTATGCGGAGATGCTACTGCCACTGCTGCTGTCC
TCGCTGCTGGGCGGGTCCCAGGCTATGGATGGGAGATTCTGGATACGAGTGCAGGAGTCAGTGATGGTGCCGGA
GGGCCTGTGCATCTCTGTGCCCTGCTCTTTCTCCTACCCCCGACAAGACTGGACAGGGTCTACCCCAGCTTATG
GCTACTGGTTCAAAGCAGTGACTGAGACAACCAAGGGTGCTCCTGTGGCCACAAACCACCAGAGTCGAGAGGTG
GAAATGAGCACCCGGGGCCGATTCCAGCTCACTGGGGATCCCGCCAAGGGGAACTGCTCCTTGGTGATCAGAGA
CGCGCAGATGCAGGATGAGTCACAGTACTTCTTTCGGGTGGAGAGAGGAAGCTATGTGAGATATAATTTCATGA
ACGATGGGTTCTTTCTAAAAGTAACAGTGCTCAGCTTCACGCCCAGACCCCAGGACCACAACACCGACCTCACC
TGCCATGTGGACTTCTCCAGAAAGGGTGTGAGCGCACAGAGGACCGTCCGACTCCGTGTGGCCTATGCCCCCAG
AGACCTTGTTATCAGCATTTCACGTGACAACACGCCAGCCCTGGAGCCCCAGCCCCAGGGAAATGTCCCATACC
TGGAAGCCCAAAAAGGCCAGTTCCTGCGGCTCCTCTGTGCTGCTGACAGCCAGCCCCCTGCCACACTGAGCTGG
GTCCTGCAGAACAGAGTCCTCTCCTCGTCCCATCCCTGGGGCCCTAGACCCCTGGGGCTGGAGCTGCCCGGGGT
GAAGGCTGGGGATTCAGGGCGCTACACCTGCCGAGCGGAGAACAGGCTTGGCTCCCAGCAGCGAGCCCTGGACC
TCTCTGTGCAGTATCCTCCAGAGAACCTGAGAGTGATGGTTTCCCAAGCAAACAGGACAGTCCTGGAAAACCTT
GGGAACGGCACGTCTCTCCCAGTACTGGAGGGCCAAAGCCTGTGCCTGGTCTGTGTCACACACAGCAGCCCCCC

AGCCAGGCTGAGCTGGACCCAGAGGGGACAGGTTCTGAGCCCCTCCCAGCCCTCAGACCCCGGGGTCCTGGAGC
TGCCTCGGGTTCAAGTGGAGCACGAAGGAGAGTTCACCTGCCACGCTCGGCACCCACTGGGCTCCCAGCACGTC
TCTCTCAGCCTCTCCGTGCACTACTCCCCGAAGCTGCTGGGCCCCTCCTGCTCCTGGGAGGCTGAGGGTCTGCA
CTGCAGCTGCTCCTCCCAGGCCAGCCCGGCCCCCTCTCTGCGCTGGTGGCTTGGGGAGGAGCTGCTGGAGGGGA
ACAGCAGCCAGGACTCCTTCGAGGTCACCCCCAGCTCAGCCGGGCCCTGGGCCAACAGCTCCCTGAGCCTCCAT
GGAGGGCTCAGCTCCGGCCTCAGGCTCCGCTGTGAGGCCTGGAACGTCCATGGGGCCCAGAGTGGATCCATCCT
GCAGCTGCCAGATAAGAAGGGACTCATCTCAACGGCATTCTCCAACGGAGCGTTTCTGGGAATCGGCATCACGG
CTCTTCTTTTCCTCTGCCTGGCCCTGATCATCATGAAGATTCTACCGAAGAGACGGACTCAGACAGAAACCCCG
AGGCCCAGGTTCTCCCGGCACAGCACGATCCTGGATTACATCAATGTGGTCCCGACGGCTGGCCCCCTGGCTCA
GAAGCGGAATCAGAAAGCCACACCAAACAGTCCTCGGACCCCTCTTCCACCAGGTGCTCCCTCCCCAGAATCAA
AGAAGAACCAGAAAAAGCAGTATCAGTTGCCCAGTTTCCCAGAACCCAAATCATCCACTCAAGCCCCAGAATCC
CAGGAGAGCCAAGAGGAGCTCCATTATGCCACGCTCAACTTCCCAGGCGTCAGACCCAGGCCTGAGGCCCGGAT
GCCCAAGGGCACCCAGGCGGATTATGCAGAAGTCAAGTTCCAATGAGGGTCTCTTAGGCTTTAGGACTGGGACT
TCGGCTAGGGAGGAAGGTAGAGTAAGAGGTTGAAGATAACAGAGTGCAAAGTTTCCTTCTCTCCCTCTCTCTCT
CTCTTTCTCTCTCTCTCTCTTTCTCTCTCTTTAAAAAACATCTGGCCAGGCAGCACAGTGGCTCACGCCTGT
AATCCCAGCACTTTGGGAGGTTGAGGTGGGCAGATCGCCTGAGGTCGGGAGTTCGAGACCAGCCTGGCCAACTT
GGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGGCGCCTGTAATCCTACCTACT
TGGGAAGCTGAGGCAGGAGAATCACTTGAACCTGGGAGACGGAGGTTGCAGTGAGCCAAGATCACACCATTGCA
CGCCAGCCTGGGCAACAAAGCGAGACTCCATCTCAAAAAAAAAATCCTCCAAATGGGTTGGGTGTCTGTAATCC
CAGCACTTTGGGAGGCTAAGGTGGGTGGATTGCTTGAGCCCAGGAGTTCGAGACCAGCCTGGGCAACATGGTGA
AACCCCATCTCTACAAAAAATACAAAACATAGCTGGGCTTGGTGGTGTGTGCCTGTAAGTCCCCAGCTGTCAGA
CATTTAAACCAGAGCAACTCCATCXTGGAATAGGAGGCTGAATAAAATGAGGCTGAGACCTXACTGGGCTGCAT
TCTCCAGACAGTGGAGGCATTCTAAGTCACAGGATGAGACAGGAGGTCCCGTACAAGATACAGGTCATAAAGAC
TTTGCTGATAAAACAGATTGCAGTAAAGAAGCCAACCAAATCCCACCAAAACCAAGTTGGCCACGAGAGTGACC
TCTGGTCGTCCTCACTGCTACACTCCTGACAGCACCATGACAGTTTACAAATGCCATGGCAACATCAGGAAGTT
ACCCGATATGTCCCAAAAGGGGGAGGAATGAATAATCCACCCCTTGTTTAGCAAATAAGCAAGAAATAACCATA
AAAGTGGGCAACCAGCAGCTCTAGGCGCTGCTCTTGTCCTATGGAGTAGCCTTCTTTTTTCCTTTAACTTTCTT
GATAAACTTGCTTTCACCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID NO:4

MLLPLLLSSLLGGSQAMDGRFWIRVQESVMVPEGLCISVPCSFSYPRQDWTGSTPAYGYWFKAVTETTKGAPVA
TNHQSREVEMSTRGRFQLTGDPAKGNCSLVIRDAQMQDESQYFFRVERGSYVRYNFMNDGFFLKVTVLSFTPRP
QDHNTDLTCHVDFSRKGVSAQRTVRLRVAYAPRDLVISISRDNTPALEPQPQGNVPYLEAQKGQFLRLLCAADS
QPPATLSWVLQNRVLSSSHPWGPRPLGLELPGVKAGDSGRYTCRAENRLGSQQRALDLSVQYPPENLRVMVSQA
NRTVLENLGNGTSLPVLEGQSLCLVCVTHSSPPARLSWTQRGQVLSPSQPSDPGVLELPRVQVEHEGEFTCHAR
HPLGSPARLSQPLRALLPEAAGPLLLLGG*

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham Corporation

(ii) TITLE OF THE INVENTION: Sialoadhesin Family 4 cDNA

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Ratner & Prestia
(B) STREET: P.O. Box 980
(C) CITY: Valley Forge
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19482

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: TO BE ASSIGNED
(B) FILING DATE: 26-MAY-1998
(C) CLASSIFICATION: UNKNOWN

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/047,572
(B) FILING DATE: 27-MAY-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Prestia, Paul F
(B) REGISTRATION NUMBER: 23,031
(C) REFERENCE/DOCKET NUMBER: GH-70026

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 610-407-0700
(B) TELEFAX: 610-407-0701
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3099 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TCCGAGAGAA CAGGCCTGTC TCAGGCAGGC CCTGCGCCTC CTATGCGGAG ATGCTACTGC   60
CACTGCTGCT GTCCTCGCTG CTGGGCGGGT CCCAGGCTAT GGATGGGAGA TTCTGGATAC  120
GAGTGCAGGA GTCAGTGATG GTGCCGGAGG GCCTGTGCAT CTCTGTGCCC TGCTCTTTCT  180
CCTACCCCCG ACAAGACTGG ACAGGGTCTA CCCCAGCTTA TGGCTACTGG TTCAAAGCAG  240
TGACTGAGAC AACCAAGGGT GCTCCTGTGG CCACAAACCA CCAGAGTCGA GAGGTGGAAA  300
TGAGCACCCG GGGCCGATTC CAGCTCACTG GGGATCCCGC CAAGGGGAAC TGCTCCTTGG  360
TGATCAGAGA CGCGCAGATG CAGGATGAGT CACAGTACTT CTTTCGGGTG GAGAGAGGAA  420
GCTATGTGAG ATATAATTTC ATGAACGATG GGTTCTTTCT AAAAGTAACA GTGCTCAGCT  480
TCACGCCCAG ACCCCAGGAC CACAACACCG ACCTCACCTG CCATGTGGAC TTCTCCAGAA  540
AGGGTGTGAG CGCACAGAGG ACCGTCCGAC TCCGTGTGGC CTATGCCCCC AGAGACCTTG  600
TTATCAGCAT TTCACGTGAC AACACGCCAG CCCTGGAGCC CCAGCCCCAG GGAAATGTCC  660
CATACCTGGA AGCCCAAAAA GGCCAGTTCC TGCGGCTCCT CTGTGCTGCT GACAGCCAGC  720
CCCCTGCCAC ACTGAGCTGG GTCCTGCAGA ACAGAGTCCT CTCCTCGTCC CATCCCTGGG  780
GCCCTAGACC CCTGGGGCTG GAGCTGCCCG GGGTGAAGGC TGGGGATTCA GGGCGCTACA  840
CCTGCCGAGC GGAGAACAGG CTTGGCTCCC AGCAGCGAGC CCTGGACCTC TCTGTGCAGT  900
ATCCTCCAGA GAACCTGAGA GTGATGGTTT CCCAAGCAAA CAGGACAGTC CTGGAAAACC  960
TTGGGAACGG CACGTCTCTC CCAGTACTGG AGGGCCAAAG CCTGTGCCTG GTCTGTGTCA 1020
CACACAGCAG CCCCCCAGCC AGGCTGAGCT GGACCCAGAG GGGACAGGTT CTGAGCCCCT 1080
CCCAGCCCTC AGACCCCGGG GTCCTGGAGC TGCCTCGGGT TCAAGTGGAG CACGAAGGAG 1140
AGTTCACCTG CCACGCTCGG CACCCACTGG GCTCCCAGCA CGTCTCTCTC AGCCTCTCCG 1200
TGCACTACTC CCCGAAGCTG CTGGGCCCCT CCTGCTCCTG GGAGGCTGAG GGTCTGCACT 1260
GCAGCTGCTC CTCCCAGGCC AGCCCGGCCC CCTCTCTGCG CTGGTGGCTT GGGGAGGAGC 1320
TGCTGGAGGG GAACAGCAGC CAGGACTCCT TCGAGGTCAC CCCCAGCTCA GCCGGGCCCT 1380
GGGCCAACAG CTCCCTGAGC CTCCATGGAG GGCTCAGCTC CGGCCTCAGG CTCCGCTGTG 1440
AGGCCTGGAA CGTCCATGGG GCCCAGAGTG GATCCATCCT GCAGCTGCCA GATAAGAAGG 1500
GACTCATCTC AACGGCATTC TCCAACGGAG CGTTTCTGGG AATCGGCATC ACGGCTCTTC 1560
TTTTCCTCTG CCTGGCCCTG ATCATCATGA AGATTCTACC GAAGAGACGG ACTCAGACAG 1620
AAACCCCGAG GCCCAGGTTC TCCCGGCACA GCACGATCCT GGATTACATC AATGTGGTCC 1680
CGACGGCTGG CCCCCTGGCT CAGAAGCGGA ATCAGAAAGC CACACCAAAC AGTCCTCGGA 1740
CCCCTCTTCC ACCAGGTGCT CCCTCCCCAG AATCAAAGAA GAACCAGAAA AAGCAGTATC 1800
AGTTGCCCAG TTTCCCAGAA CCCAAATCAT CCACTCAAGC CCCAGAATCC CAGGAGAGCC 1860
AAGAGGAGCT CCATTATGCC ACGCTCAACT TCCCAGGCGT CAGACCCAGG CCTGAGGCCC 1920
GGATGCCCAA GGGCACCCAG GCGGATTATG CAGAAGTCAA GTTCCAATGA GGGTCTCTTA 1980
GGCTTTAGGA CTGGGACTTC GGCTAGGGAG GAAGGTAGAG TAAGAGGTTG AAGATAACAG 2040
AGTGCAAAGT TTCCTTCTCT CCCTCTCTCT CTCTCTTTCT CTCTCTCTCT CTCTTTCTCT 2100
CTCTTTAAAA AACATCTGGC CAGGCAGCAC AGTGGCTCAC GCCTGTAATC CCAGCACTTT 2160
GGGAGGTTGA GGTGGGCAGA TCGCCTGAGG TCGGGAGTTC GAGACCAGCC TGGCCAACTT 2220
GGTGAAACCC CGTCTCTACT AAAAATACAA AAATTAGCTG GGCATGGTGG CAGGCGCCTG 2280
TAATCCTACC TACTTGGGAA GCTGAGGCAG GAGAATCACT TGAACCTGGG AGACGGAGGT 2340
TGCAGTGAGC CAAGATCACA CCATTGCACG CCAGCCTGGG CAACAAAGCG AGACTCCATC 2400
TCAAAAAAAA AATCCTCCAA ATGGGTTGGG TGTCTGTAAT CCCAGCACTT TGGGAGGCTA 2460
AGGTGGGTGG ATTGCTTGAG CCCAGGAGTT CGAGACCAGC CTGGGCAACA TGGTGAAACC 2520
CCATCTCTAC AAAAAATACA AAACATAGCT GGGCTTGGTG GTGTGTGCCT GTAAGTCCCC 2580
AGCTGTCAGA CATTTAAACC AGAGCAACTC CATCXTGGAA TAGGAGGCTG AATAAAATGA 2640
GGCTGAGACC TXACTGGGCT GCATTCTCCA GACAGTGGAG GCATTCTAAG TCACAGGATG 2700
AGACAGGAGG TCCCGTACAA GATACAGGTC ATAAAGACTT TGCTGATAAA ACAGATTGCA 2760
GTAAAGAAGC CAACCAAATC CCACCAAAAC CAAGTTGGCC ACGAGAGTGA CCTCTGGTCG 2820
TCCTCACTGC TACACTCCTG ACAGCACCAT GACAGTTTAC AAATGCCATG GCAACATCAG 2880
GAAGTTACCC GATATGTCCC AAAAGGGGGA GGAATGAATA ATCCACCCCT TGTTTAGCAA 2940
ATAAGCAAGA AATAACCATA AAAGTGGGCA ACCAGCAGCT CTAGGCGCTG CTCTTGTCCT 3000
ATGGAGTAGC CTTCTTTTTT CCTTTAACTT TCTTGATAAA CTTGCTTTCA CCTAAAAAAA 3060
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA                        3099
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 639 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Leu Leu Pro Leu Leu Leu Ser Ser Leu Leu Gly Gly Ser Gln Ala
 1               5                  10                  15
Met Asp Gly Arg Phe Trp Ile Arg Val Gln Glu Ser Val Met Val Pro
            20                  25                  30
Glu Gly Leu Cys Ile Ser Val Pro Cys Ser Phe Ser Tyr Pro Arg Gln
        35                  40                  45
Asp Trp Thr Gly Ser Thr Pro Ala Tyr Gly Tyr Trp Phe Lys Ala Val
    50                  55                  60
Thr Glu Thr Thr Lys Gly Ala Pro Val Ala Thr Asn His Gln Ser Arg
65                  70                  75                  80
Glu Val Glu Met Ser Thr Arg Gly Arg Phe Gln Leu Thr Gly Asp Pro
                85                  90                  95
Ala Lys Gly Asn Cys Ser Leu Val Ile Arg Asp Ala Gln Met Gln Asp
            100                 105                 110
Glu Ser Gln Tyr Phe Phe Arg Val Glu Arg Gly Ser Tyr Val Arg Tyr
        115                 120                 125
Asn Phe Met Asn Asp Gly Phe Phe Leu Lys Val Thr Val Leu Ser Phe
    130                 135                 140
Thr Pro Arg Pro Gln Asp His Asn Thr Asp Leu Thr Cys His Val Asp
145                 150                 155                 160
Phe Ser Arg Lys Gly Val Ser Ala Gln Arg Thr Val Arg Leu Arg Val
                165                 170                 175
Ala Tyr Ala Pro Arg Asp Leu Val Ile Ser Ile Ser Arg Asp Asn Thr
            180                 185                 190
Pro Ala Leu Glu Pro Gln Pro Gln Gly Asn Val Pro Tyr Leu Glu Ala
        195                 200                 205
Gln Lys Gly Gln Phe Leu Arg Leu Leu Cys Ala Ala Asp Ser Gln Pro
    210                 215                 220
Pro Ala Thr Leu Ser Trp Val Leu Gln Asn Arg Val Leu Ser Ser Ser
225                 230                 235                 240
His Pro Trp Gly Pro Arg Pro Leu Gly Leu Glu Leu Pro Gly Val Lys
                245                 250                 255
Ala Gly Asp Ser Gly Arg Tyr Thr Cys Arg Ala Glu Asn Arg Leu Gly
            260                 265                 270
Ser Gln Gln Arg Ala Leu Asp Leu Ser Val Gln Tyr Pro Pro Glu Asn
        275                 280                 285
Leu Arg Val Met Val Ser Gln Ala Asn Arg Thr Val Leu Glu Asn Leu
    290                 295                 300
Gly Asn Gly Thr Ser Leu Pro Val Leu Glu Gly Gln Ser Leu Cys Leu
305                 310                 315                 320
Val Cys Val Thr His Ser Ser Pro Pro Ala Arg Leu Ser Trp Thr Gln
                325                 330                 335
Arg Gly Gln Val Leu Ser Pro Ser Gln Pro Ser Asp Pro Gly Val Leu
            340                 345                 350
Glu Leu Pro Arg Val Gln Val Glu His Glu Gly Glu Phe Thr Cys His
```

```
        355                 360                 365
Ala Arg His Pro Leu Gly Ser Gln His Val Ser Leu Ser Leu Ser Val
    370                 375                 380
His Tyr Ser Pro Lys Leu Leu Gly Pro Ser Cys Ser Trp Glu Ala Glu
385                 390                 395                 400
Gly Leu His Cys Ser Cys Ser Ser Gln Ala Ser Pro Ala Pro Ser Leu
                405                 410                 415
Arg Trp Trp Leu Gly Glu Glu Leu Leu Glu Gly Asn Ser Ser Gln Asp
            420                 425                 430
Ser Phe Glu Val Thr Pro Ser Ser Ala Gly Pro Trp Ala Asn Ser Ser
        435                 440                 445
Leu Ser Leu His Gly Gly Leu Ser Ser Gly Leu Arg Leu Arg Cys Glu
    450                 455                 460
Ala Trp Asn Val His Gly Ala Gln Ser Gly Ser Ile Leu Gln Leu Pro
465                 470                 475                 480
Asp Lys Lys Gly Leu Ile Ser Thr Ala Phe Ser Asn Gly Ala Phe Leu
                485                 490                 495
Gly Ile Gly Ile Thr Ala Leu Leu Phe Leu Cys Leu Ala Leu Ile Ile
            500                 505                 510
Met Lys Ile Leu Pro Lys Arg Arg Thr Gln Thr Glu Thr Pro Arg Pro
        515                 520                 525
Arg Phe Ser Arg His Ser Thr Ile Leu Asp Tyr Ile Asn Val Val Pro
    530                 535                 540
Thr Ala Gly Pro Leu Ala Gln Lys Arg Asn Gln Lys Ala Thr Pro Asn
545                 550                 555                 560
Ser Pro Arg Thr Pro Leu Pro Pro Gly Ala Pro Ser Pro Glu Ser Lys
                565                 570                 575
Lys Asn Gln Lys Lys Gln Tyr Gln Leu Pro Ser Phe Pro Glu Pro Lys
            580                 585                 590
Ser Ser Thr Gln Ala Pro Glu Ser Gln Glu Ser Gln Glu Glu Leu His
        595                 600                 605
Tyr Ala Thr Leu Asn Phe Pro Gly Val Arg Pro Arg Pro Glu Ala Arg
    610                 615                 620
Met Pro Lys Gly Thr Gln Ala Asp Tyr Ala Glu Val Lys Phe Gln
625                 630                 635
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 3099 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TCCGAGAGAA CAGGCCTGTC TCAGGCAGGC CCTGCGCCTC CTATGCGGAG ATGCTACTGC    60
CACTGCTGCT GTCCTCGCTG CTGGGCGGGT CCCAGGCTAT GGATGGGAGA TTCTGGATAC   120
GAGTGCAGGA GTCAGTGATG GTGCCGGAGG GCCTGTGCAT CTCTGTGCCC TGCTCTTTCT   180
CCTACCCCCG ACAAGACTGG ACAGGGTCTA CCCCAGCTTA TGGCTACTGG TTCAAAGCAG   240
TGACTGAGAC AACCAAGGGT GCTCCTGTGG CCACAAACCA CCAGAGTCGA GAGGTGGAAA   300
TGAGCACCCG GGGCCGATTC CAGCTCACTG GGGATCCCGC CAAGGGGAAC TGCTCCTTGG   360
TGATCAGAGA CGCGCAGATG CAGGATGAGT CACAGTACTT CTTTCGGGTG GAGAGAGGAA   420
GCTATGTGAG ATATAATTTC ATGAACGATG GGTTCTTTCT AAAAGTAACA GTGCTCAGCT   480
TCACGCCCAG ACCCCAGGAC CACAACACCG ACCTCACCTG CCATGTGGAC TTCTCCAGAA   540
```

```
AGGGTGTGAG CGCACAGAGG ACCGTCCGAC TCCGTGTGGC CTATGCCCCC AGAGACCTTG    600
TTATCAGCAT TTCACGTGAC AACACGCCAG CCCTGGAGCC CCAGCCCCAG GGAAATGTCC    660
CATACCTGGA AGCCCAAAAA GGCCAGTTCC TGCGGCTCCT CTGTGCTGCT GACAGCCAGC    720
CCCCTGCCAC ACTGAGCTGG GTCCTGCAGA ACAGAGTCCT CTCCTCGTCC CATCCCTGGG    780
GCCCTAGACC CCTGGGGCTG GAGCTGCCCG GGGTGAAGGC TGGGGATTCA GGGCGCTACA    840
CCTGCCGAGC GGAGAACAGG CTTGGCTCCC AGCAGCGAGC CCTGGACCTC TCTGTGCAGT    900
ATCCTCCAGA GAACCTGAGA GTGATGGTTT CCCAAGCAAA CAGGACAGTC CTGGAAAACC    960
TTGGGAACGG CACGTCTCTC CCAGTACTGG AGGGCCAAAG CCTGTGCCTG GTCTGTGTCA   1020
CACACAGCAG CCCCCCAGCC AGGCTGAGCT GGACCCAGAG GGGACAGGTT CTGAGCCCCT   1080
CCCAGCCCTC AGACCCCGGG GTCCTGGAGC TGCCTCGGGT TCAAGTGGAG CACGAAGGAG   1140
AGTTCACCTG CCACGCTCGG CACCCACTGG GCTCCCAGCA CGTCTCTCTC AGCCTCTCCG   1200
TGCACTACTC CCCGAAGCTG CTGGGCCCCT CCTGCTCCTG GGAGGCTGAG GGTCTGCACT   1260
GCAGCTGCTC CTCCCAGGCC AGCCCGGCCC CCTCTCTGCG CTGGTGGCTT GGGGAGGAGC   1320
TGCTGGAGGG GAACAGCAGC CAGGACTCCT TCGAGGTCAC CCCCAGCTCA GCCGGGCCCT   1380
GGGCCAACAG CTCCCTGAGC CTCCATGGAG GGCTCAGCTC CGGCCTCAGG CTCCGCTGTG   1440
AGGCCTGGAA CGTCCATGGG GCCCAGAGTG GATCCATCCT GCAGCTGCCA GATAAGAAGG   1500
GACTCATCTC AACGGCATTC TCCAACGGAG CGTTTCTGGG AATCGGCATC ACGGCTCTTC   1560
TTTTCCTCTG CCTGGCCCTG ATCATCATGA AGATTCTACC GAAGAGACGG ACTCAGACAG   1620
AAACCCCGAG GCCCAGGTTC TCCCGGCACA GCACGATCCT GGATTACATC AATGTGGTCC   1680
CGACGGCTGG CCCCCTGGCT CAGAAGCGGA ATCAGAAAGC CACACCAAAC AGTCCTCGGA   1740
CCCCTCTTCC ACCAGGTGCT CCCTCCCCAG AATCAAAGAA GAACCAGAAA AAGCAGTATC   1800
AGTTGCCCAG TTTCCCAGAA CCCAAATCAT CCACTCAAGC CCCAGAATCC CAGGAGAGCC   1860
AAGAGGAGCT CCATTATGCC ACGCTCAACT TCCCAGGCGT CAGACCCAGG CCTGAGGCCC   1920
GGATGCCCAA GGGCACCCAG GCGGATTATG CAGAAGTCAA GTTCCAATGA GGGTCTCTTA   1980
GGCTTTAGGA CTGGGACTTC GGCTAGGGAG GAAGGTAGAG TAAGAGGTTG AAGATAACAG   2040
AGTGCAAAGT TTCCTTCTCT CCCTCTCTCT CTCTCTTTCT CTCTCTCTCT CTCTTTCTCT   2100
CTCTTTAAAA AACATCTGGC CAGGCAGCAC AGTGGCTCAC GCCTGTAATC CCAGCACTTT   2160
GGGAGGTTGA GGTGGGCAGA TCGCCTGAGG TCGGGAGTTC GAGACCAGCC TGGCCAACTT   2220
GGTGAAACCC CGTCTCTACT AAAAATACAA AAATTAGCTG GGCATGGTGG CAGGCGCCTG   2280
TAATCCTACC TACTTGGGAA GCTGAGGCAG GAGAATCACT TGAACCTGGG AGACGGAGGT   2340
TGCAGTGAGC CAAGATCACA CCATTGCACG CCAGCCTGGG CAACAAAGCG AGACTCCATC   2400
TCAAAAAAAA AATCCTCCAA ATGGGTTGGG TGTCTGTAAT CCCAGCACTT TGGGAGGCTA   2460
AGGTGGGTGG ATTGCTTGAG CCCAGGAGTT CGAGACCAGC CTGGGCAACA TGGTGAAACC   2520
CCATCTCTAC AAAAAATACA AAACATAGCT GGGCTTGGTG GTGTGTGCCT GTAAGTCCCC   2580
AGCTGTCAGA CATTTAAACC AGAGCAACTC CATCXTGGAA TAGGAGGCTG AATAAAATGA   2640
GGCTGAGACC TXACTGGGCT GCATTCTCCA GACAGTGGAG GCATTCTAAG TCACAGGATG   2700
AGACAGGAGG TCCCGTACAA GATACAGGTC ATAAAGACTT TGCTGATAAA ACAGATTGCA   2760
GTAAAGAAGC CAACCAAATC CCACCAAAAC CAAGTTGGCC ACGAGAGTGA CCTCTGGTCG   2820
TCCTCACTGC TACACTCCTG ACAGCACCAT GACAGTTTAC AAATGCCATG GCAACATCAG   2880
GAAGTTACCC GATATGTCCC AAAAGGGGGA GGAATGAATA ATCCACCCCT TGTTTAGCAA   2940
ATAAGCAAGA AATAACCATA AAAGTGGGCA ACCAGCAGCT CTAGGCGCTG CTCTTGTCCT   3000
ATGGAGTAGC CTTCTTTTTT CCTTTAACTT TCTTGATAAA CTTGCTTTCA CCTAAAAAAA   3060
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA                          3099
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Leu Pro Leu Leu Leu Ser Ser Leu Leu Gly Gly Ser Gln Ala

```
1               5                  10                  15
Met Asp Gly Arg Phe Trp Ile Arg Val Gln Glu Ser Val Met Val Pro
            20                  25                  30
Glu Gly Leu Cys Ile Ser Val Pro Cys Ser Phe Ser Tyr Pro Arg Gln
        35                  40                  45
Asp Trp Thr Gly Ser Thr Pro Ala Tyr Gly Tyr Trp Phe Lys Ala Val
    50                  55                  60
Thr Glu Thr Thr Lys Gly Ala Pro Val Ala Thr Asn His Gln Ser Arg
65                  70                  75                  80
Glu Val Glu Met Ser Thr Arg Gly Arg Phe Gln Leu Thr Gly Asp Pro
                85                  90                  95
Ala Lys Gly Asn Cys Ser Leu Val Ile Arg Asp Ala Gln Met Gln Asp
            100                 105                 110
Glu Ser Gln Tyr Phe Phe Arg Val Glu Arg Gly Ser Tyr Val Arg Tyr
        115                 120                 125
Asn Phe Met Asn Asp Gly Phe Phe Leu Lys Val Thr Val Leu Ser Phe
    130                 135                 140
Thr Pro Arg Pro Gln Asp His Asn Thr Asp Leu Thr Cys His Val Asp
145                 150                 155                 160
Phe Ser Arg Lys Gly Val Ser Ala Gln Arg Thr Val Arg Leu Arg Val
                165                 170                 175
Ala Tyr Ala Pro Arg Asp Leu Val Ile Ser Ile Ser Arg Asp Asn Thr
            180                 185                 190
Pro Ala Leu Glu Pro Gln Pro Gln Gly Asn Val Pro Tyr Leu Glu Ala
        195                 200                 205
Gln Lys Gly Gln Phe Leu Arg Leu Leu Cys Ala Ala Asp Ser Gln Pro
    210                 215                 220
Pro Ala Thr Leu Ser Trp Val Leu Gln Asn Arg Val Leu Ser Ser Ser
225                 230                 235                 240
His Pro Trp Gly Pro Arg Pro Leu Gly Leu Glu Leu Pro Gly Val Lys
                245                 250                 255
Ala Gly Asp Ser Gly Arg Tyr Thr Cys Arg Ala Glu Asn Arg Leu Gly
            260                 265                 270
Ser Gln Gln Arg Ala Leu Asp Leu Ser Val Gln Tyr Pro Pro Glu Asn
        275                 280                 285
Leu Arg Val Met Val Ser Gln Ala Asn Arg Thr Val Leu Glu Asn Leu
    290                 295                 300
Gly Asn Gly Thr Ser Leu Pro Val Leu Glu Gly Gln Ser Leu Cys Leu
305                 310                 315                 320
Val Cys Val Thr His Ser Ser Pro Pro Ala Arg Leu Ser Trp Thr Gln
                325                 330                 335
Arg Gly Gln Val Leu Ser Pro Ser Gln Pro Ser Asp Pro Gly Val Leu
            340                 345                 350
Glu Leu Pro Arg Val Gln Val Glu His Glu Gly Glu Phe Thr Cys His
        355                 360                 365
Ala Arg His Pro Leu Gly Ser Pro Ala Arg Leu Ser Gln Pro Leu Arg
    370                 375                 380
Ala Leu Leu Pro Glu Ala Ala Gly Pro Leu Leu Leu Leu Gly Gly
385                 390                 395
```

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

(i) an isolated polypeptide comprising an amino acid sequence selected from the group having at least:

(a) 70% identity;

(b) 80% identity;

(c) 90% identity; or

(d) 95% identity

to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2;

(ii) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2 or

(iii) an isolated polypeptide which is the amino acid sequence of SEQ ID NO:2.

2. An isolated polynucleotide selected from the group consisting of:

(i) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least

(a) 70% identity;

(b) 80% identity;

(c) 90% identity; or

(d) 95% identity;

to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2;

(ii) an isolated polynucleotide comprising a nucleotide sequence that has at least:

(a) 70% identity

(b) 80% identity;

(c) 90% identity; or

(d) 95% identity;

over its entire length to a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(iii) an isolated polynucleotide comprising a nucleotide sequence which has at least:

(a) 70% identity;

(b) 80% identity;

(c) 90% identity; or

(d) 95% identity;

to that of SEQ ID NO: 1 over the entire length of SEQ ID NO:1;

(iv) an isolated polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(vi) an isolated polynucleotide which is the polynucleotide of SEQ ID NO: 1; or

(vi) an isolated polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof.;

or a nucleotide sequence complementary to said isolated polynucleotide.

3. An antibody immunospecific for the polypeptide of claim 1.

4. A method for the treatment of a subject:

(i) in need of enhanced activity or expression of the polypeptide of claim 1 comprising:

(a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or

(b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence encoding said polypeptide in a form so as to effect production of said polypeptide activity *in vivo*.; or

(ii) having need to inhibit activity or expression of the polypeptide of claim 1 comprising:

(a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or

(b) administering to the subject a nucleic acid molecule that inhibits the expression of a nucleotide sequence encoding said polypeptide; and/or

(c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said polypeptide for its ligand, substrate , or receptor.

5. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the polypeptide of claim 1 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or

(b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

6. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presence of a labeled competitor;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

7. An agonist or an antagonist of the polypeptide of claim 1.

8. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.

9. A process for producing a recombinant host cell comprising transforming or transfecting a cell with the expression system of claim 8 such that the host cell, under appropriate culture conditions, produces a polypeptide comprising an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2.

10. A recombinant host cell produced by the process of claim 9.

11. A membrane of a recombinant host cell of claim 10 expressing a polypeptide comprising an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2.

12. A process for producing a polypeptide comprising culturing a host cell of claim 10 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.

13. An isolated polynucleotide selected form the group consisting of:

(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 70%, 80%, 90%, 95%, 97% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3;

(c) the polynucleotide of SEQ ID NO:3; or

(d) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 70%, 80%, 90%, 95%, 97-99% identity to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4.

14. A polypeptide selected from the group consisting of:

(a) a polypeptide which comprises an amino acid sequence which has at least 70%, 80%, 90%, 95%, 97-99% identity to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) a polypeptide which has an amino acid sequence which is at least 70%, 80%, 90%, 95%, 97-99% identity to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(c) a polypeptide which comprises the amino acid of SEQ ID NO:4;

(d) a polypeptide which is the polypeptide of SEQ ID NO:4;

(e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/10791

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 143.1; 435/69.1, 71.1, 440, 455, 471; 514/8, 44; 530/350, 387.1, 388.2, 388.22; 536/23.1, 24.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG, BIOSIS, CA, EMBASE, MEDLINE, WPI
search terms: sialoadhesin, saf-4, cd33, SEQ ID NOS. 1-4

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CROCKER et al. Sialoadhesin, a macrophage sialic acid binding receptor for haemopoieticel cells with 17 immunoglobulin-like domains. EMBO J. 1994, Vol. 13, No. 19, pages 4490-4503, see entire document.	1-14
Y	MUCKLOW et al. Sialoadhesin (Sn) maps to mouse chromosome 2 and human chromosome 20 and is not linked to the other members of the sialoadhesin family, CD22, MAG and CD33. Genomics. 1995, Vol. 28, pages 344-346, see entire document.	1-14
Y	FREEMAN et al. Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules, Blood. 15 April 1995, Vol. 85, No. 8, pages 2005-2012, see entire document.	1-14

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 AUGUST 1998	30 SEP 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Deborah Lawrence For PHILLIP GAMBEL Telephone No. (703) 308-0196

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A61K 38/00, 39/395, 48/00; C07H 21/04; C07K 14/435, 14/705, 16/00, 16/18, 16/28; C12N 15/01, 15/11, 15/12, 15/63

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/130.1, 143.1; 435/69.1, 71.1, 440, 455, 471; 514/8, 44; 530/350, 387.1, 388.2, 388.22; 536/23.1, 24.5